

0570
1/27

45

OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/710,633

DATE: 11/21/2001

TIME: 12:26:17

Input Set : N:\Crf3\RULE60\09710633.raw

Output Set: N:\CRF3\11212001\I710633.raw

1 <110> APPLICANT: Kent, Stephen B.H.
2 Muir, Tom W.
3 Dawson, Philip E.
4 <120> TITLE OF INVENTION: SYNTHESIS OF PROTEIN BY NATIVE CHEMICAL LIGATION
5 <130> FILE REFERENCE: gry0030p
6 <140> CURRENT APPLICATION NUMBER: 09/710,633
7 <141> CURRENT FILING DATE: 2000-11-08
8 <150> PRIOR APPLICATION NUMBER: 08/945,997
9 <151> PRIOR FILING DATE: 1998-02-12
10 <160> NUMBER OF SEQ ID NOS: 20
11 <170> SOFTWARE: PatentIn Ver. 2.1
13 <210> SEQ ID NO: 1
14 <211> LENGTH: 5
15 <212> TYPE: PRT
16 <213> ORGANISM: Artificial Sequence
17 <220> FEATURE:
18 <223> OTHER INFORMATION: Description of Artificial Sequence: peptide
19 <221> NAME/KEY: SITE
20 <222> LOCATION: (5)
21 <223> OTHER INFORMATION: wherein COSH is thioacid
22 <400> SEQUENCE: 1
23 Leu Tyr Arg Ala Gly
24 1 5
26 <210> SEQ ID NO: 2
27 <211> LENGTH: 6
28 <212> TYPE: PRT
29 <213> ORGANISM: Artificial Sequence
30 <220> FEATURE:
31 <223> OTHER INFORMATION: Description of Artificial Sequence: peptide
32 <400> SEQUENCE: 2
33 Cys Arg Ala Glu Tyr Ser
34 1 5
36 <210> SEQ ID NO: 3
37 <211> LENGTH: 5
38 <212> TYPE: PRT
39 <213> ORGANISM: Artificial Sequence
40 <220> FEATURE:
41 <223> OTHER INFORMATION: Description of Artificial Sequence: peptide
42 <221> NAME/KEY: SITE
43 <222> LOCATION: (5)
44 <223> OTHER INFORMATION: wherein COSBn is benzyl thioester
45 <400> SEQUENCE: 3
46 Leu Tyr Arg Ala Gly
47 1 5
49 <210> SEQ ID NO: 4
50 <211> LENGTH: 5
51 <212> TYPE: PRT

ENTERED

RAW SEQUENCE LISTING

DATE: 11/21/2001

PATENT APPLICATION: US/09/710,633

TIME: 12:26:17

Input Set : N:\Crf3\RULE60\09710633.raw

Output Set: N:\CRF3\11212001\I710633.raw

52 <213> ORGANISM: Artificial Sequence
53 <220> FEATURE:
54 <223> OTHER INFORMATION: Description of Artificial Sequence: peptide
55 <221> NAME/KEY: SITE
56 <222> LOCATION: (5)
57 <223> OTHER INFORMATION: wherein Gly is modified and represented by Gly-
58 alphaCOS-CH2C(NHAc)CO2H
59 <400> SEQUENCE: 4
60 Leu Tyr Arg Ala Gly
61 1 5
63 <210> SEQ ID NO: 5
64 <211> LENGTH: 11
65 <212> TYPE: PRT
66 <213> ORGANISM: Artificial Sequence
67 <220> FEATURE:
68 <223> OTHER INFORMATION: Description of Artificial Sequence: peptide
69 <400> SEQUENCE: 5
70 Leu Tyr Arg Ala Gly Cys Arg Ala Glu Tyr Ser
71 1 5 10
73 <210> SEQ ID NO: 6
74 <211> LENGTH: 5
75 <212> TYPE: PRT
76 <213> ORGANISM: Artificial Sequence
77 <220> FEATURE:
78 <223> OTHER INFORMATION: Description of Artificial Sequence: peptide
79 <221> NAME/KEY: SITE
80 <222> LOCATION: (5)
81 <223> OTHER INFORMATION: wherein SCH2COOH is 2-thioacetic acid
82 <400> SEQUENCE: 6
83 Leu Tyr Arg Ala Gly
84 1 5
86 <210> SEQ ID NO: 7
87 <211> LENGTH: 33
88 <212> TYPE: PRT
89 <213> ORGANISM: Homo sapiens
90 <220> FEATURE:
91 <221> NAME/KEY: SITE
92 <222> LOCATION: (33)
93 <223> OTHER INFORMATION: wherein COSH is thioacid
94 <221> NAME/KEY: SITE
95 <222> LOCATION: (1)
96 <223> OTHER INFORMATION: wherein Msc is 2-methyl-sulfonyl-ethyloxy-carbonyl
97 <400> SEQUENCE: 7
98 Ser Ala Lys Glu Leu Arg Cys Gln Cys Ile Lys Thr Tyr Ser Lys Pro
99 1 5 10 15
100 Phe His Pro Lys Phe Ile Lys Glu Leu Arg Val Ile Glu Ser Gly Pro
101 20 25 30
102 Ala
104 <210> SEQ ID NO: 8

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/710,633

DATE: 11/21/2001

TIME: 12:26:17

Input Set : N:\Crf3\RULE60\09710633.raw

Output Set: N:\CRF3\11212001\I710633.raw

```

105 <211> LENGTH: 33
106 <212> TYPE: PRT
107 <213> ORGANISM: Homo sapiens
108 <220> FEATURE:
109 <221> NAME/KEY: SITE
110 <222> LOCATION: (33)
111 <223> OTHER INFORMATION: wherein COSBn is benzyl thioester
112 <400> SEQUENCE: 8
113   Ser Ala Lys Glu Leu Arg Cys Gln Cys Ile Lys Thr Tyr Ser Lys Pro
114       1           5           10           15
115   Phe His Pro Lys Phe Ile Lys Glu Leu Arg Val Ile Glu Ser Gly Pro
116           20           25           30
117   Ala
119 <210> SEQ ID NO: 9
120 <211> LENGTH: 39
121 <212> TYPE: PRT
122 <213> ORGANISM: Homo sapiens
123 <400> SEQUENCE: 9
124   Cys Ala Asn Thr Glu Ile Ile Val Lys Leu Ser Asp Gly Arg Glu Leu
125       1           5           10           15
126   Cys Leu Asp Pro Lys Glu Asn Trp Val Gln Arg Val Val Glu Lys Phe
127           20           25           30
128   Leu Lys Arg Ala Glu Asn Ser
129           35
131 <210> SEQ ID NO: 10
132 <211> LENGTH: 72
133 <212> TYPE: PRT
134 <213> ORGANISM: Homo sapiens
135 <220> FEATURE:
136 <221> NAME/KEY: SITE
137 <222> LOCATION: (72)
138 <223> OTHER INFORMATION: SH4
139 <400> SEQUENCE: 10
140   Ser Ala Lys Glu Leu Arg Cys Gln Cys Ile Lys Thr Tyr Ser Lys Pro
141       1           5           10           15
142   Phe His Pro Lys Phe Ile Lys Glu Leu Arg Val Ile Glu Ser Gly Pro
143           20           25           30
144   Ala Cys Ala Asn Thr Glu Ile Ile Val Lys Leu Ser Asp Gly Arg Glu
145           35           40           45
146   Leu Cys Leu Asp Pro Lys Glu Asn Trp Val Gln Arg Val Val Glu Lys
147           50           55           60
148   Phe Leu Lys Arg Ala Glu Asn Ser
149           65           70
151 <210> SEQ ID NO: 11
152 <211> LENGTH: 40
153 <212> TYPE: PRT
154 <213> ORGANISM: Human immunodeficiency virus
155 <220> FEATURE:
156 <221> NAME/KEY: SITE

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/710,633

DATE: 11/21/2001

TIME: 12:26:17

Input Set : N:\Crf3\RULE60\09710633.raw

Output Set: N:\CRF3\11212001\I710633.raw

157 <222> LOCATION: (40)
158 <223> OTHER INFORMATION: wherein COSNB is 5-thio-2-nitro-benzoic acid ester
159 <400> SEQUENCE: 11
160 Pro Gln Ile Thr Leu Trp Lys Arg Pro Leu Val Thr Ile Arg Ile Gly
161 1 5 10 15
162 Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
163 20 25 30
164 Ile Glu Glu Met Asn Leu Pro Gly
165 35 40
167 <210> SEQ ID NO: 12
168 <211> LENGTH: 59
169 <212> TYPE: PRT
170 <213> ORGANISM: Human immunodeficiency virus
171 <220> FEATURE:
172 <221> NAME/KEY: SITE
173 <222> LOCATION: (27)
174 <223> OTHER INFORMATION: wherein Xaa is 2-Aminobutyric acid
175 <221> NAME/KEY: SITE
176 <222> LOCATION: (55)
177 <223> OTHER INFORMATION: wherein Xaa is 2-Aminobutyric acid
178 <400> SEQUENCE: 12
179 Cys Trp Lys Pro Lys Met Ile Gly Gly Ile Gly Gly Phe Ile Lys Val
180 1 5 10 15
W--> 181 Arg Gln Tyr Asp Gln Ile Pro Val Glu Ile Xaa Gly His Lys Ala Ile
182 20 25 30
183 Gly Thr Val Leu Val Gly Pro Thr Pro Val Asn Ile Ile Gly Arg Asn
184 35 40 45
W--> 185 Leu Leu Thr Gln Ile Gly Xaa Thr Leu Asn Phe
186 50 55
188 <210> SEQ ID NO: 13
189 <211> LENGTH: 40
190 <212> TYPE: PRT
191 <213> ORGANISM: Human immunodeficiency virus
192 <220> FEATURE:
193 <221> NAME/KEY: SITE
194 <222> LOCATION: (40)
195 <223> OTHER INFORMATION: wherein COSBn is ??
196 <221> NAME/KEY: SITE
197 <222> LOCATION: (40)
198 <223> OTHER INFORMATION: wherein COSBn is benzyl thio ester
199 <400> SEQUENCE: 13
200 Pro Gln Ile Thr Leu Trp Lys Arg Pro Leu Val Thr Ile Arg Ile Gly
201 1 5 10 15
202 Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
203 20 25 30
204 Ile Glu Glu Met Asn Leu Pro Gly
205 35 40
207 <210> SEQ ID NO: 14
208 <211> LENGTH: 40

RAW SEQUENCE LISTING

DATE: 11/21/2001

PATENT APPLICATION: US/09/710,633

TIME: 12:26:17

Input Set : N:\Crf3\RULE60\09710633.raw

Output Set: N:\CRF3\11212001\I710633.raw

209 <212> TYPE: PRT
210 <213> ORGANISM: Human immunodeficiency virus
211 <220> FEATURE:
212 <221> NAME/KEY: SITE
213 <222> LOCATION: (40)
214 <223> OTHER INFORMATION: wherein COSPh is phenyl thioester
215 <400> SEQUENCE: 14
216 Pro Gln Ile Thr Leu Trp Lys Arg Pro Leu Val Thr Ile Arg Ile Gly
217 1 5 10 15
218 Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
219 20 25 30
220 Ile Glu Glu Met Asn Leu Pro Gly
221 35 40
223 <210> SEQ ID NO: 15
224 <211> LENGTH: 99
225 <212> TYPE: PRT
226 <213> ORGANISM: Human immunodeficiency virus
227 <220> FEATURE:
228 <221> NAME/KEY: SITE
229 <222> LOCATION: (67)
230 <223> OTHER INFORMATION: wherein Xaa is amino butyric acid
231 <221> NAME/KEY: SITE
232 <222> LOCATION: (95)
233 <223> OTHER INFORMATION: wherein Xaa is 2-Aminobutyric acid
234 <400> SEQUENCE: 15
235 Pro Gln Ile Thr Leu Trp Lys Arg Pro Leu Val Thr Ile Arg Ile Gly
236 1 5 10 15
237 Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
238 20 25 30
239 Ile Glu Glu Met Asn Leu Pro Gly Cys Trp Lys Pro Lys Met Ile Gly
240 35 40 45
241 Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Val
242 50 55 60
W--> 243 Glu Ile Xaa Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
244 65 70 75 80
W--> 245 Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Xaa Thr
246 85 90 95
247 Leu Asn Phe
249 <210> SEQ ID NO: 16
250 <211> LENGTH: 48
251 <212> TYPE: PRT
252 <213> ORGANISM: Bacillus amyloliquefaciens
253 <220> FEATURE:
254 <221> NAME/KEY: SITE
255 <222> LOCATION: (48)
256 <223> OTHER INFORMATION: wherein COSNB is 5-thio-2-nitro benzoic acid ester
257 <400> SEQUENCE: 16
258 Ala Gln Val Ile Asn Thr Phe Asp Gly Val Ala Asp Tyr Leu Gln Thr
259 1 5 10 15

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/710,633

DATE: 11/21/2001

TIME: 12:26:18

Input Set : N:\Crf3\RULE60\09710633.raw

Output Set: N:\CRF3\11212001\I710633.raw

L:181 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:12
L:185 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:12
L:243 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:15
L:245 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:15